Machine Learning Over Networks, EP3620

Grading Assignment (peer reviews)

Group 1: Yasaman, Firooz, Ozan

March 21, 2023

1 HW1: Graded by Yasaman and Firooz (solutions of group 2)

Question 1:
1- correct
2- correct
3a- not correct, the inequality directions are different in 4 and 5. Moreover, instead of the "applying norm" term, it is better to call it applying Cauchy-Schwartz inequality.
3b- correct
3c- correct
3d- correct

Question 2:
a- correct
b- correct
c- correct

Question 3:
a- correct
b- correct
c- correct
d- correct

Question 4:
a- correct
b- correct

c- not correct, the hessian of f is diagonal so finding its inverse is not the problem.
d- correct (again for part c you should consider the main issue)

Question 5: The approach is correct. However, for complete proof, you need to provide more details on your assumptions. For example, I think you should prove why $\phi(x)$ is $(L-\mu)$-smooth and then after that proof you can use the lemma that says it has a coercive gradient.

2 HW2: graded by Ozan (solutions of group 3)

1- correct
2- correct
3- In the expression between (1) and (2), there should be expectation taken for the f(w) function so that we can go from (1) to (2). However following steps are correct.

3 HW3: graded by Firooz (solutions of group 4)

Question 1:
That is a nice approach which was completely correct.

Question 2:
Thanks for proving the first two theorems in detail. The answer is correct, but I am not sure about the power $2k$. You did not answer whether the solution is primal feasible or not.

Question 3:
The answer is correct. You have to compare your derived algorithm with the primal method regarding the total number of communications as well.

4 CA1: graded by Firooz (solutions of group 2)

a- correct
b- correct
c- correct
d- correct

5 CA2: graded by Ozan (solutions of group 3)

Overall, the results are reasonable and the comparisons are correct. There are two minor issues:

- In the given data, y values do not have to be 1 or -1. It is correct that in logistic regression thats the general case, but not in this dataset.
- If the computation time or memory results had been given, the last part (the comparison of algorithms) could be based on these results rather than a general comparison of algorithms.

6 CA3: graded by Yasaman (solutions of group 4)

a- I checked your backpropagation code and also the GD, SGD and etc algorithms which were totally correct. My only concern is your first figure where SGD has better performance than GD (lower loss) which I believe is not correct. In fact, GD should have the lowest loss. Besides, I think the figure that plots loss versus times should be increasing and the GD method is the most time-consuming approach (which in this figure we could not see it).